



Академия АйТи
a Softline Company



Потоковая обработка данных с Apache Flink

Код курса: A-FLINK

Потоковая обработка данных с Apache Flink

Код курса: A-FLINK

Длительность	24 ак. часа
Формат	
Разработчик курса	Академия АйТи
Тип	Учебный курс
Способ обучения	Под руководством тренера

О курсе

Курс по Apache Flink обучает потоковой обработке данных в реальном времени, используя DataStream API и Flink SQL. Участники освоят создание и оптимизацию решений для обработки потоков данных, управление состоянием, обработку временных меток, агрегацию данных с окнами и интеграцию с внешними системами (Kafka, HDFS). Курс подходит для инженеров данных и разработчиков, работающих с большими данными и стремящихся улучшить навыки в области потоковой аналитики и реального времени.

Подробная информация

Профиль аудитории:

- разработчики Hadoop
- разработчики Spark
- разработчики ETL
- специалисты по аналитике и архитекторов данных
- архитекторы, инженеры и разработчики больших данных

Предварительные требования:

- основы Python или Java
- основы потоковой обработки данных
- знания о распределённых системах
- знания SQL: Полезны для работы с Flink SQL

По окончании курса слушатели смогут:

- проектировать и разрабатывать системы потоковой обработки данных с использованием Apache Flink
- использовать DataStream API для обработки и агрегации данных в реальном времени
- применять методы управления состоянием и обеспечения отказоустойчивости
- разрабатывать приложения для аналитики в реальном времени с использованием Flink SQL и Table API

- работать с экосистемой коннекторов для интеграции с различными источниками данных

Программа курса

Модуль 1: Введение в потоковую обработку и Apache Flink

- Введение в обработку данных в реальном времени.
- Архитектура потоковой обработки данных.
- Зачем использовать Apache Flink: преимущества перед другими решениями (например, Apache Spark).
- Ключевые компоненты Flink и обзор экосистемы.

Модуль 2: Основы DataStream API

- Введение в DataStream API: работа с потоками данных.
- Трансформации данных: фильтрация, агрегирование, трансформация.
- Типы преобразований и их применение в реальных задачах.

Модуль 3: Обработка состояний и управление состоянием

- Стейтфул потоки: понятие состояний в потоках.
- Как поддерживать состояние в Flink и решать проблемы с состоянием.
- Реализация контроля состояний с использованием Managed States и Checkpoints.

Модуль 4: Время событий и Watermarks

- Работа с временем в потоках: Event Time vs Processing Time.
- Что такое Watermarks и как их использовать для обработки задержанных данных.
- Настройка обработок поздних элементов и их использование для повышения точности обработки.

Модуль 5: Окна в потоковой аналитике

- Использование окон в потоках: Tumbling, Sliding, Session Windows.
- Применение окон для агрегации данных в реальных сценариях аналитики.

Модуль 6: Подключения и экосистема коннекторов Flink

- Введение в экосистему коннекторов Flink.
- Подключение Flink с внешними системами: Kafka, HDFS, JDBC.

Модуль 7: Введение в Flink SQL и Table API

- Основы работы с Flink SQL: что это и как используется.
- Табличный API в Flink: SQL-подход для обработки данных.

Модуль 8: Примеры использования и паттерны приложений

- Применение Flink для построения аналитических приложений в реальном времени.
- Примеры использования Flink в event-driven приложениях, обработке данных и построении ETL-

пайплайнов.

[Посмотреть расписание курса и записаться на обучение](#)

Обращайтесь по любым вопросам
к менеджерам Академии АйТи

+7 (495) 150 96 00 | academy@academyit.ru